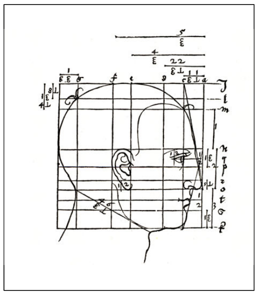
Metodologija psiholoških istraživanja
1

obrada faktorijalni 1



26. novembar 2018.

IV. Obrada podataka
B. Faktorijalni nacrti
1. Jednofaktorski nacrti
(a) organizacija podataka
(b) deskriptivne statističke mere
mere prebrojavanja
mere centralne tendencije
mere varijabilnosti
(c) standardne mere
(d) transformacije skupova mera

B. Obrada faktorijalnih nacrta
2

- podsetimo se:
- obrada podataka za tri vrste nacrta:
  - *frekvencijski nacrti:*
    - *3 tipa:* nacrti sa jednom, sa dve i sa tri varijable
  - *faktorijalni nacrti:*
    - *3 tipa:* nacrti sa dve, sa tri i sa četiri varijable
  - *korelaciono-regresioni nacrti:*
    - *2 tipa:* nacrti sa dve i sa tri varijable

Faktorijalni nacrti
- nacrti sa kategoričkim NV i numeričkim ZV
- bavićemo se nacrtima sa:
  - 1, 2, i 3 NV (faktori)
  - 1 ZV

B. Obrada faktorijalnih nacrta
3

1. Jednofaktorski nacrti (JFN)
- faktorijalni nacrti sa jednom NV (faktorom) i jednom ZV
  - *broj nivoa faktora*: dva (*bivalentan* faktor), tri (*trivalentan* faktor), četiri, itd
  - *ponovljenost faktora (nacrta):* mogu biti neponovljeni i ponovljeni
  - *podaci:* vrednosti NV i ZV za svaki objekt
  - *deskriptivna statistika:* proseci ZV za svaki nivo (kategoriju) faktora (NV)
  - *statistika zaključivanja:* postojanje stat. znač. razlika među prosecima
- *PRIMERI:* bivalentni faktori
  - neponovljeni nacrt: visine muškaraca i žena
    - *faktor (NV):* pol; *nivoi faktora:* muški pol, ženski pol
    - *ZV:* visina subjekata
    - *istraživačko pitanje:* da li postoji razlika u proseku visina dva pola?
  - ponovljeni nacrt: vreme reakcije na dve boje
    - *faktor (NV):* boja; *nivoi faktora:* crvena, zelena
    - *ZV:* vreme reakcije
    - *istraživačko pitanje:* da li postoji razlika u proseku vremena reakcije na boje?

1. Jednofaktorski nacrti (JFN)
4

a. Organizacija podataka
- matrice podataka: dva tipa
- (1) matrica tipa objekti x varijable (OxV)
  - kao kod frekvencijskih nacrta
  - *redovi:* objekti; *kolone:* varijable
    - u JFN: *dve* kolone, za NV i za ZV
  - pogodno za *neponovljene* nacrte
    - npr. pol i visina
- (2) matrica tipa objekti x nivoi (OxN)
  - novi oblik matrica
  - *redovi:* objekti; *kolone:* nivoi NV (faktora)
    - ima onoliko kolona koliko NV ima nivoa
    - za bivalentni nacrt: dve kolone; za trival. nacrt: 3 kolone, ...
  - pogodno i za neponovljene i za ponovljene nacrte
    - npr. pol i visina; vreme reakcije na boje
- *važno:* programi za statističku obradu podataka obično ’očekuju’ da u istom redu ’nađu’ podatke koji se odnose na isti objekt, što znači:
  - *neponovljeni nacrti:* koristiti matrice OxV; *ponovljeni nacrti:* koristiti matrice OxN

objekti x varijable

#	POL	VISINA
1.	1	180
2.	1	185
3.	1	170
4.	1	190
5.	1	175
6.	2	165
7.	2	150
8.	2	160
9.	2	170
10	2	155

neponovljen nacrt

objekti x nivoi

#	muš.	žene
1.	180	165
2.	185	150
3.	170	160
4.	190	170
5.	175	155

neponovljen nacrt

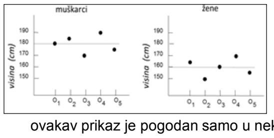
#	Crv.	Zel.
1.	200	195
2.	150	200
3.	250	205
4.	220	190
5.	180	210

ponovljen nacrt

1. Jednofaktorski nacrti (JFN)
5

- prikaz podataka
  - osim dva matrična prikaza podataka, u JFN koriste se i *grafički* prikazi
  - za razliku od frekvencijskih nacrta, kod kojih nismo razmatrali grafičke prikaze *podataka* već samo *rezultata* (frekvenci, procenta, proporcija)
- primeri grafičkih prikaza podataka

(a) 2D prikaz: *x-osa:* oznake objekata, *y-osa:* mere ZV

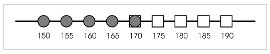


pojedinačni podaci: prikazani *markerima* (oznakama) objekata

proseci ZV: prikazani horizontalnim linijama

ovakav prikaz je pogodan samo u nekim slučajevima

(b) prikaz na brojnoj osi (x-osa): mere i markeri objekata

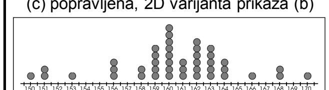


mana prikaza: za objekte sa istom merom markeri se preklapaju

ovakav prikaz uglavnom nije pogodan, ali se može popraviti

1. Jednofaktorski nacrti (JFN)
6

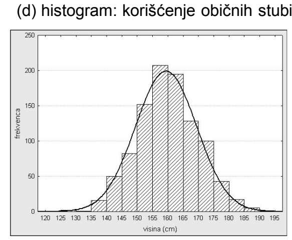
(c) popravljena, 2D varijanta prikaza (b)



markeri objekata sa *istom* merom ZV postavljaju se jedni *iznad* drugih, čineći *markerske stubiće*

visina stubića izražava *frekvencu* mere

(d) histogram: korišćenje običnih stubića umesto markerskih



daleko najčešći način prikaza podataka

mere ZV na x-osi su obično grupisane u *intervale* (razrede)

visina stubića (štapića) odgovara *broju* markera u prikazu (c), tj. markerskim stubićima

histogram prikazuje *distribuciju* (raspodelu) podataka

distribucija se često može aproksimirati *normalnom (Gausovom)* krivom

x-osa: mere ZV, *y-osa:* frekvenca

1. Jednofaktorski nacrti (JFN) 7

b. Deskriptivne statističke mere: tri grupe mera

mere prebrojavanja, mere centralne tendencije, mere varijabilnosti

(1) mere prebrojavanja: odnose se na *broj podataka* u istraživanju

- zavisno od broja podataka na pojedinim *nivoima* faktora, nacrt može biti:
  - balansiran: ako je broj podataka jednak na svim nivoima
  - nebalansiran: ako je broj podataka različit na nekim nivoima
- *PRIMER:* istraživanje zavisnosti vremena reakcije od rukosti (dva nivoa)
  - nebalansiran nacrt: 10 levorukih i 100 desnorukih
    - problem: podaci na različitim nivoima nisu jednako pouzdani
  - balansiran nacrt: 100 levorukih i 100 desnorukih; 10 levorukih i 10 desnorukih
    - pogodniji za obradu i interpretaciju
- uočiti: faktorijalni i frekvencijski nacrti se *razlikuju* u tretiranju broja podataka
- *PRIMER: varijabla:* rukost; *kategorije:* levoruki i desnoruki
- *frekvencijski nacrt: cilj:* kolike su frekvence kategorija (lev. i des.) u uzorku
  - eksperimentator *ne* (želi da) kontroliše frekvence, već ih utvrđuje empirijski
- *faktorijalni nacrt: cilj:* kako ZV (vreme reakcije) zavisi od NV (rukosti)
  - eksperimentator *može* (i treba) da kontroliše frekvence (zbog balansiranja)
  - eksperimentator empirijski utvrđuje *proseke* ZV na različitim nivoima NV

1. Jednofaktorski nacrti (JFN) 8

(2) mere centralne tendencije

- kada se utvrde vrednosti neke kvantitativne varijable dobijaju se:
  - *(individualne) mere, (individualne) opservacije, skalari*
  - sve ovakve mere uzete zajedno čine: *skup mera, varijablu, vektor*
- osnovna ideja mera centralne tendencije:
  - definisati jedan broj koji karakteriše ceo skup mera
  - taj jedan broj treba da bude *reprezentativan* odn. *tipičan* za ceo skup brojeva, da izražava *tendenciju* celog skupa, da bude neka vrsta *centra* tog skupa
- ova osnovna ideja se može ***matematički*** uobličiti na tri načina
  - naime, za meru centralne tendencije može se uzeti ili mera koja je u datom skupu *najčešća*, ili mera koja je *srednja*, ili mera koja je *prosečna*
- (1) mod (engleski: 'mode')
  - definicija: *najčešća* mera u datom skupu mera
  - *PRIMERI:* kao primere za mere cent. tend. koristićemo dva mala skupa mera
    - prvi skup: {0, 1, 3, 3, 8}; za ovaj skup mod je 3 (jer trojki ima najviše)
    - drugi skup: {0, 1, 1, 3, 3, 8}; u ovom skupu postoje dva moda, 1 i 3
  - mod je jednostavna mera, ali se ne koristi često kao mera centralne tendencije, jer uglavnom nije mnogo informativna za skup

1. Jednofaktorski nacrti (JFN) 9

- (2) medijana (engleski: 'median')
- definicija: *srednja* mera u datom skupu mera
  - izračunavanje veličine medijane vrši se u dva koraka:
    - prvo se dati skup mera *poređa* u niz po veličini
    - zatim se traži broj koji je u tom nizu na *sredini* odn. u *centru*
  - budući da je medijana u centru, *jednak* broj mera je manji i veći od medijane
  - zavisno od toga da li u skupu ima neparan ili paran broj mera, medijana se različito računa
- *PRIMERI:* ista dva skupa kao za mod
  - za skup {0, 1, 3, 3, 8} medijana je 3
    - kad je broj mera neparan, medijana je mera koja je tačno na sredini
    - u skupu ima 5 mera, a 3 je treća, srednja mera (dve su pre med. i dve posle)
  - za skup {0, 1, 1, 3, 3, 8} medijana je 2
    - u skupu ima paran broj mera (6), pa ne postoji jedinstvena srednja mera, već dve, a to su 1 (treća mera) i 3 (četvrta mera)
    - medijana se računa kao *prosek* dve srednje mere: (1+3)/2 = 2
      - u skupu su tri mere pre medijane (0, 1, 1), a tri su posle (3, 3, 8)
- medijana se koristi znatno češće od moda, ali znatno ređe nego prosek

1. Jednofaktorski nacrti (JFN) 10

- (3) aritmetička sredina (prosek) (engleski: '(arithmetic) mean')
- definicija: *prosečna* mera u datom skupu mera
  - u matematici postoje i srodne ali drugačije mere, *geometrijska sredina* i *harmonijska sredina*, ali se one ne koriste mnogo u statistici
  - prosek skupa mera označava se često sa M (od 'mean')
  - *individualne mere* datog skupa mera označavaćemo sa Y, a njihov zbir sa ΣY
- izračunavanje proseka skupa mera vrši se u tri koraka:
  - *1. saberu* se individualne mere, i dobije se njihov zbir: ΣY
  - *2. prebroje* se individualne mere, i dobije se njihov broj: N
  - 3. dobijeni zbir se *podeli* sa dobijenim brojem: $M = \Sigma Y / N$
  - *PRIMERI:* ista dva skupa, {0, 1, 3, 3, 8} i {0, 1, 1, 3, 3, 8}
    - prvi skup: zbir: $\Sigma Y=0+1+3+3+8=15$; broj: N=5; prosek: $M = \Sigma Y / N = 15/5 = 3$
    - drugi skup: zbir: $\Sigma Y=0+1+1+3+3+6=16$; broj: N=6; prosek: $M = \Sigma Y / N = 16/6 = 2.666...$
- korisno je utvrditi i *devijacije* (odstupanja) individualnih mera od proseka
  - oznaka: d; račun: prema tzv. devijacionoj jednačini: $d = Y – M$
  - za prvi skup mera, {0, 1, 3, 3, 8 }, skup devijacija od proseka 3 je {-3, -2, 0, 0, 5 }
    - naime: -3 = 0-3, -2 = 1-3, 0 = 3-3, 0 = 3-3, 5 = 8-3
  - za drugi skup mera, {0, 1, 1, 3, 3, 8 }, skup dev. je {-2.67,-1.67,-1.67,0.33,0.33,5.33}
  - uočiti: zbir devijacija je nula (i mora da bude nula!)

1. Jednofaktorski nacrti (JFN) 11

- veoma odstupajuće mere od proseka se zovu *autlajeri* (engleski: 'outlier')
  - mogući srpski termini: 'izuzetak', 'štrčak', 'iznimak'
- ovakve mere mogu odražavati *stvarnu* raznolikost podataka date varijable, ali mogu i *nepripadati* ispitivanoj populaciji, ili čak biti *greške* u zapisu
- veličina *moda* i *medijane* obično ne zavisi mnogo od veoma odstupajućih mera, ali veličina *proseka* može znatno da zavisi od njih
- *PRIMER:* poredimo mere centr. tendencije skupa {0, 1, 3, 3, 8} i {0, 1, 3, 3, 88}
  - *mod:* i za prvi i za drugi skup mod je 3
  - *medijana:* i za prvi i za drugi skup medijana je 3
  - *prosek:* za prvi skup prosek je 3, a za drugi skup prosek je 19!
- kaže se: prosek je manje *robustna* mera od moda i medijane
  - naime: prosek je manje *otporan* na autlajere, više se *menja* nego mod i medijana
- robustnost proseka zavisi od *veličine* datog skupa
  - što je skup veći, uticaj veličine pojedinačnih mera na prosek je manji
- mada se prosek najviše koristi, za neke skupove medijana je pogodnija
  - PRIMER: prosek i medijana dužine studiranja; visine primanja
    - postoji veoma mali broj studenata koji studiraju veoma dugo, i stoga 'veštački' podižu prosek dužine studiranja, ali ne utiču mnogo na medijanu

1. Jednofaktorski nacrti (JFN) 12

#	muš.	žene
1.	180	165
2.	185	150
3.	170	160
4.	190	170
5.	175	155

- u JFN postoje dve vrste proseka, *grupni* i *opšti*
- *PRIMER:* visine muškaraca (grupa 1) i žena (grupa 2)
- *grupni proseci:* proseci mera *pojedinih nivoa* (odn. grupa)
  - podaci za grupu 1 (muškarci): {180, 185, 170, 190, 175}
    - *oznake:* individ. mere: Y1; broj mera: N1; prosek: M1
    - *račun*: $\Sigma Y1 = 180+...+175 = 900$; $N1 = 5$; $M1 = \Sigma Y1/N1 = 900/5 = 180$
  - podaci za grupu 2 (žene): {165, 150, 160, 170, 155}
    - *oznake:* individualne mere: Y2; broj mera: N2; prosek: M2
    - *račun*: $\Sigma Y2 = 165+...+155 = 800$; $N2 = 5$; $M2 = \Sigma Y2/N2 = 800/5 = 160$
  - kada je nacrt *balansiran*, tj. N1 = N2, često se oba broja označavaju sa N
- *opšti (totalni) prosek:* prosek mera *celog uzorka*
  - oznake: individualne mere: Y; broj mera: N1+N2 = N+N = 2N; prosek: My
  - podaci za obe grupe, tj. sve članove uzorka: {180, 185, ... ,170, 155}
    - važi: $\Sigma Y = \Sigma Y1+\Sigma Y2=1700$; $2N = 10$; $My = \Sigma Y/2N = 1700/10 = 170$
- u *balansiranim* nacrtima, opšti prosek je *prosek* grupnih proseka, tj. njihova sredina:
  - $My = (M1+M2)/2 = (180+160)/2 = 170$
- u *nebalansiranim* nacrtima, to ne mora biti slučaj, npr. opšti prosek vremena reakcija 10 levorukih i 100 desnorukih je mnogo bliži proseku desnorukih nego levorukih

1. Jednofaktorski nacrti (JFN) 13

- u JFN postoje dve vrste devijacija, *grupne* i *opšte (totalne)*
- *PRIMER:* visine muškaraca (grupa 1) i žena (grupa 2)
- *grupne devijacije:* odstupanja individualnih mera od proseka *svojih grupa*
  - grupa 1 (muškarci): d1 = Y1 – M1
    - mere: {180, 185, 170, 190, 175}, M1=180
    - devijacije od 180: {0, +5, -10, +10, -5}
    - uočimo: Σd1 = 0 + (+5) + (-10) + (+10) + (-5) = 0
  - grupa 2 (žene): d2 = Y2 – M2
    - mere {165, 150, 160, 170, 155}, M2=160
    - devijacije od M2=160: {+5, -10, 0, +10, -5}
    - uočimo: Σd2 = +5 + (-10) + (0) + (+10) + (-5) = 0 (pa je i prosek 0)
- *opšte (totalne) devijacije:* odstupanja ind. mera od proseka *celog uzorka*
  - dy = Y – My
  - mere: {180, 185, 170, 190, 175, 165, 150, 160, 170, 155}, My=170
  - devijacije od My=170: {+10, +15, 0, +20, +5, -5, -20, -10, 0, -15 }
  - uočimo: Σdy = 10 +15 + ... -15 = 0
- napomena: ako nije važna razlika između grupnih i opštih devijacija, sve devijacije označavamo sa d, individualne mere sa Y, a proseke sa M, prema formuli d = Y - M

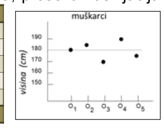
#	muš.	žene
1.	180	165
2.	185	150
3.	170	160
4.	190	170
5.	175	155
M	180	160

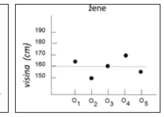
1. Jednofaktorski nacrti (JFN) 14

- uočimo: bilo da su grupne ili opšte, pojedinačne devijacije mogu biti:
  - pozitivne (d>0): to je slučaj ako je individualna mera *veća* od proseka (Y>M)
  - negativne (d<0): to je slučaj ako je individualna mera *manja* od proseka (Y<M)
  - nulte (d=0): to je slučaj ako je individualna mera *jednaka* proseku (Y=M)
- numerički i grafički prikaz podataka, proseka i devijacija

#	muš.	žene
1.	180	165
2.	185	150
3.	170	160
4.	190	170
5.	175	155
M	180	160

#	muš. dev.	žene dev.
1.	0	+5
2.	+5	-10
3.	-10	0
4.	+10	+10
5.	-5	-5
M	0	0





- pojedinačni podaci Y1 i Y2: prikazani *markerima* (oznakama) na dijagramima
- proseci M1 i M2: prikazani *horizontalnim* linijama na visini proseka ('linije proseka')
- devijacije d1 i d2: prikazane *odstojanjima* markera od linija proseka
  - Y>M, d>0: marker iznad linije;Y<M, d<0: marker ispod linije; Y=M, d=0: marker na liniji
- linije proseka imaju osobinu da su *centralne*, tj. da prolaze *sredinom* skupa markera
  - linije su *centralne:* u smislu da je *zbir* odstojanja markera *iznad* linija proseka jednak zbiru odstojanja markera *ispod* linija proseka
  - ova osobina sledi iz činjenice da je zbir svih devijacija nula

1. Jednofaktorski nacrti (JFN) 15

- strukturna jednačina
  - iz *devijacione jednačine* (koja glasi, prema definiciji devijacije): d = Y – M
  - sledi *strukturna jednačina:* Y = M + d
    - *PRIMER:* Y = 190, M =180, d = 10
      - devijaciona jednačina: d = Y – M odnosno 10 = 190 - 180
      - strukturna jednačina: Y = M + d odnosno 190 = 180 + 10
- strukturna jednačina prikazuje *strukturu* individualnih mera
  - individualna mera Y se sastoji od dve komponente, M i d:
    - prva komponenta, prosek M, je *zajednička* svim merama u skupu
    - druga komponenta, devijacija d, je *specifična* za svaku pojed. meru u skupu
- jedna interpretacija proseka i devijacija
  - proseci se, naročito u tehnici i prirodnim naukama, koriste prilikom obrade podataka *višestrukih merenja* iste pojave (na pr. položaja neke zvezde na nebu)
  - usled neminovnih grešaka prilikom merenja, pojedinačne mere se *razlikuju* jedne od drugih, a prosek daje *pouzdaniju* meru od svih pojedinačnih mera
  - u takvim situacijama, prosek skupa, M, se može smatrati kao najbolja *pretpostavka* (*teorija, model*) o stvarnoj meri neke pojave
    - stoga se proseci ponekad nazivaju *očekivane mere*

1. Jednofaktorski nacrti (JFN) 16

- pojedinačne mere, Y, odstupaju od modela, tj. proseka M, usled dejstva raznih nesistematskih spoljnih varijabli
- sa ovakve tačke gledišta, devijacije, d, odražavaju *nesavršenost* modela, njegov neuspeh da potpuno tačno predvidi pojedinačna merenja
- strukturna jednačina Y = M + d se može opisno i metaforički izraziti na više načina:
  - *podatak* = *model* + *odstupanje*
  - *dato merenje* = *tačna mera* + *greška merenja*
  - *činjenica* = *očekivanje* + *iznenađenje*
  - *pojedinačno* = *opšte* + *posebno*
  - ...
- usled ovakvih razmatranja, odstupanja se često nazivaju *greške* (engleski: 'error')
  - međutim, ukažimo da je, iako odomaćen u statistici, termin 'greška' većinom neprikladan
  - na pr., u slučaju merenja visine, nikako nije tačno da svi subjekti ustvari imaju istu visinu, a da su odstupanja pojedinih subjekata samo greške nastale usled nesavršenosti merenja
- napomena: postoji formalna *sličnost* mera u JFN i UFN:

JFN	*skup opservir. mera:* **Y**	*zbir mera:* **ΣY**	*prosek mera:* **M = ΣY/N**	*devijacije mera:* **d = Y – M** (važi: Σd = 0)
UFN	*skup opserv. frekvenci:* **f**	*totalna frekvenca:* **N = Σf**	*očekivana frekvenca:* **f' = Σf/k**	*rezidualne frekvence:* **d = f–f'** (važi: Σd = 0)

- *razlika:* u JFN postoje *dve grupe* mera (dva nivoa), pa postoje Y1 i Y2, M1 i M2, itd

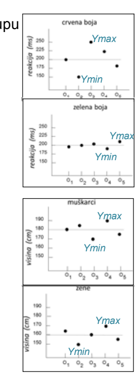
1. Jednofaktorski nacrti (JFN) 17

- (3) mere varijabilnosti
- u skupu mera neke varijable neće skoro nikad sve mere biti *iste* veličine
  - skoro uvek će postojati razlike tj. *varijabilnost* (*disperzija, raspršenost*) mera
  - varijabilnost je važan aspekt skupa mera, pored proseka
  - dva skupa mogu imati iste proseke ali različitu varijabilnost, i time se razlikovati
  - *PRIMERI: proseci su jednaki a varijabilnosti različite*
    - varijabilnost količine leka u pilulama dve kompanije
    - varijabilnost plata u različitim zemljama
- osnovna ideja mera varijabilnosti je da se definiše:
  - jedna mera koja karakteriše stepen *variranja* mera u okviru celog skupa mera
- kako definisati i konstruisati takvu meru, kako naći prikladnu formulu?
  - ključna osobina takve mere je da bude *proporcionalna* varijabilnosti, tj.:
    - da bude što *veći* broj što je varijabilnost skupa mera oko proseka veća
    - da bude *nula* ako u skupu mera nema varijabilnosti (tj. ako su sve mere iste)
    - da ne može da bude *negativna*
- postoji *više načina* da se *matematički uobliči* ova osnovna ideja
  - stoga postoji ne jedna nego *veći broj* mera varijabilnosti, a razmotrićemo šest

1. Jednofaktorski nacrti (JFN) 18

(1) Raspon: R = Ymax - Ymin

- razlika između najveće mere (Ymax) i najmanje mere (Ymin) u skupu
- *PRIMERI:*
  - vreme reakcije na crvenu i zelenu boju:
    - proseci: Mcrv = 200 ms, Mzel = 200 ms
    - varijabilnost:
      - crvena: R=Ymax – Ymin = 250 ms – 150 ms = 100 ms
      - zelena: R= Ymax – Ymin = 210 ms – 190 ms = 20 ms
    - uočiti: proseci dve grupe su jednaki, a varijabilnosti su različite
  - visine muškaraca i žena
    - proseci: Mm = 180 cm, Mž = 160 cm
    - varijabilnost:
      - muškarci: R = Ymax – Ymin = 190 – 170 = 20 cm
      - žene: R = Ymax – Ymin = 170 – 150 = 20 cm
    - uočiti: proseci dve grupe su različiti, a varijabilnosti su jednake
- raspon je korisna i intuitivno jasna mera varijabilnosti
  - međutim, raspon ne uzima u obzir *sve* mere u skupu, već samo *dve* (najveću i najmanju)
  - postoje bolje mere varijabilnosti, koje uzimaju u obzir sve mere u skupu



1. Jednofaktorski nacrti (JFN) 19

- (2) Prosečno odstupanje: PO = Σd / N
  - formula je konstruisana po analogiji sa formulom za prosek: M = ΣY / N
    - kao što je M karakteristična *mera*, tako bi PO bilo karakteristično *odstupanje*
  - *veliki problem:* PO je *uvek nula*, jer je zbir Σd = 0, pa je i prosek Σd/N nula
  - *PRIMER:* devijacije visina kod muškaraca su +5, 0, -10, +10, -5, sa zbirom 0
  - *zaključak:* PO je *beskorisna* kao mera varijabilnosti, i zato se i ne koristi
- (3) Prosečno apsolutno odstupanje: PAO = Σ|d| / N = Σ|Y-M| / N
  - *PRIMER: apsolutne* devijacije visina kod muškaraca su: 5, 0, 10, 10, 5, zbir = 30
    - PAO = Σ|d| / N = 30 / 5 = 6 cm
  - PAO odražava prosečnu devijaciju, ali bez obzira na *smer* odstupanja (tj. da li su mere manje ili veće od proseka)
  - PAO ima upravo one osobine koje se traže za meru varijabilnosti:
    - PAO će biti veća što su devijacije veće (po apsolutnoj vrednosti), tj. ukoliko je varijabilnost mera oko proseka veća
    - PAO može biti nula samo ako su sve devijacije nulte
      - tada su sve mere jednake proseku, odn. međusobno, pa ni nema varijabilnosti
    - PAO ne može biti negativno, jer se zasniva na apsolutnim vrednostima
    - zaključak: za razliku od PO, PAO je *korisna* mera varijabilnosti
  - problem: PAO ima izvesne statističke mane, pa su druge mere pogodnije

1. Jednofaktorski nacrti (JFN) 20

- (4) Varijansa: V = Σd² / N = Σ(Y-M)² / N
  - *PRIMER:* kvadrati devijacija visina kod mušk. su: 25, 0, 100, 100, 25, zbir = 250
    - V = Σd² / N = 250/5 = 50
  - slično kao što je PAO prosečno *apsolutno* odstupanje, varijansa je prosečno *kvadrirano* odstupanje (PKO)
  - kvadriranje je statistički pogodnija operacija od apsolutne vrednosti
  - kao i PAO, varijansa ima sledeće tri povoljne osobine za meru varijabilnosti:
    - utoliko je *veća* ukoliko su devijacije veće (bez obzira na predznak)
    - *nulta* je ako u podacima nema nema varijabilnosti
    - ne može biti *negativna*, jer kvadrati ne mogu biti negativni
  - problem: varijansa je, usled kvadriranja, intuitivno manje jasna mera nego PAO
- (5) Zbir kvadrata: SS = Σd² = Σ(Y-M)² (engl.: 'Sum of Squares')
  - *PRIMER:* kvadrati devijacija visina kod muškaraca su: 25, 0, 100, 100, 25
    - SS = Σd² = 250
  - SS, budući da je brojilac varijanse, ima slične tri osobine, naime:
    - utoliko je veći ukoliko su devijacije veće (bez obzira na predznak)
    - nulti je ako u podacima nema varijabilnosti; ne može biti negativan
  - kao što ćemo videti, SS se u nekim primenama koristi kao mera varijabilnosti
  - varijansa je pogodnija mera utoliko što odražava *prosek* d² (usled deljenja sa N)

1. Jednofaktorski nacrti (JFN) 21

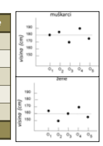
- (6) Standardna devijacija: SD = √V = √(Σd² / N) = √(Σ(Y-M)² / N)
  - *PRIMER:* visine kod muškaraca
    - SD = √V = √50 = 7.1
  - SD ima slične tri povoljne osobine kao PAO, V, i SS:
    - utoliko je veća ukoliko su devijacije veće (bez obzira na predznak)
    - nulta je ako u podacima nema varijabilnosti
    - ne može biti negativna
  - SD je u izvesnom smislu primerenija i smislenija mera nego V
    - naime, usled korenovanja, SD se (kao i PAO), izražava istim *jedinicama* kao individualne mere, a ne njihovim kvadratima
    - *PRIMER:* u primeru sa visinama, SD = 7.1 cm, dok je V = 50 cm²
  - SD ima i neke matematički pogodne osobine, vezane za normalnu krivu
  - SD je najčešće korišćena mera varijabilnosti
  - odnos SD i PAO:
    - matematički su različite; u primeru, PAO = 6 cm, SD = 7.1 cm
    - SD je manje intuitivna od PAO, usled veće matematičke složenosti
    - PAO i SD su ipak slične, utoliko što obe izražavaju *tipičnu* devijaciju
      - kod PAO ta devijacija je *prosečna*, kod SD je *standardna*

1. Jednofaktorski nacrti (JFN) 22

poređenje mera varijabilnosti grupa u dva istraživanja

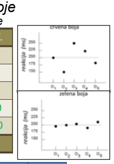
visine muškaraca i žena

	mere		devijacije	
#	muš.	žene	muš.	žene
1.	180	165	0	+5
2.	185	150	+5	-10
3.	170	160	-10	0
4.	190	170	+10	+10
5.	175	155	-5	-5
M	180	160	0	0



vreme reakcije na boje

	mere		devijacije	
#	Crv.	Zel.	Crv.	Zel.
1.	200	195	0	-5
2.	150	200	-50	0
3.	250	205	+50	+5
4.	220	190	20	-10
5.	180	210	-20	+10
M	200	200	0	0



naziv	formula	muš.	žene	crv.	zel.
Raspon	Ymax - Ymin	20	20	100	20
PO	Σd / N	0	0	0	0
PAO	Σ|d| / N	6	6	28	6
V	Σd² / N	50	50	5800	250
SS	Σd²	250	250	29000	1250
SD	√(Σd² / N)	7.1	7.1	34.1	7.1

uočiti: sve mere, osim raspona, zasnovane su na *devijacijama* d od proseka

prema svim merama: varijabilnost u obe grupe je jednaka (jer su devijacije u obe grupe iste)

prema svim merama (osim PO): veća je varijabilnost u prvoj grupi

1. Jednofaktorski nacrti (JFN) 23

- mere centralne tendencije i varijabilnosti mogu se odnositi:
  - na *uzorak*, i tada se nazivaju *statistici*, označeni sa M, SD, V itd
  - na *populaciju*, i tada se nazivaju *parametri*, označeni sa μ, σ, σ² itd
- vrednosti *statistika* se dobijaju empirijskim istraživanjima
- vrednosti *parametara* se *procenjuju* na osnovu vrednosti statistika
  - na pr., prosek populacije, μ, se procenjuje prosekom uzorka, M
    - ako se na reprezentativnom uzorku muškaraca utvrdi prosek od 180 cm (statistik M), ta vrednost je najbolja moguća procena proseka visine u populaciji (parametar μ)
- međutim, u slučaju stand. devijacije i varijanse, situacija je komplikovanija
  - može se pokazati da su statistici SD i V tzv. *pristrasne* procene parametara σ i σ²
    - grubo rečeno, ovi statistici nisu *najbolje* moguće procene odgovarajućih parametara
- u statističkoj teoriji se dokazuje da se ovi parametri mogu *bolje* proceniti ako se u formulama za V i SD umesto izraza N koristi izraz N-1, tj ako se umesto da se koristi:
  - *uzoračka* varijansa: V = Σd² / N, i *uzoračka* stand. dev.: SD = √Σd² / N, koriste tzv.
  - *populaciona* varijansa: V = Σd² / (N–1), i *populaciona* st. dev.: SD = √Σd² / (N-1)
- izraz N – 1 se odnosi na broj stepena slobode (df)
  - naime: u skupu od N mera postoji, u principu, N stepeni slobode
  - ali: ako znamo njihov *prosek*, jedan stepen slobode se izgubio, pa je df = N - 1
  - za skup od N devijacija takođe važi df = N – 1, jer se devijacije moraju sabrati do 0

1. Jednofaktorski nacrti (JFN) 24

c. standardne mere

- kad utvrdimo neki skup mera i njihov prosek M, možemo izračunati za svaku *meru* Y njoj odgovarajuću *devijaciju* d
- devijacija neke mere pruža informaciju o *odstupanju* mere od proseka, i time o *odnosu* date mere prema ostalim merama u skupu
- kada imamo informaciju o tome kolika je *mera* Y:
  - znamo: vrednost varijable za dati objekt
  - ali: ne znamo *odnos* te vrednosti prema proseku i ostalim vrednostima u skupu
  - *PRIMERI:*
    - data nam je mera koja je broj poena studenta na testu, npr. 21
      - znamo koliki je uspeh, ali ne znamo da li je 21 dobar ili loš uspeh, tj. da li je veći ili manji od proseka
    - data nam je mera koja je visina neke ženske osobe, npr. 170cm
      - znamo samu visinu, ali ne znamo da li je 170 mala ili velika visina u poređenju sa prosečnom visinom žena
- kada imamo informaciju o tome kolika je *devijacija* d:
  - ne znamo: samu vrednost varijable za dati objekt
  - ali: znamo za koliko je ta vrednost veća ili manja od proseka svih mera u skupu

1. Jednofaktorski nacrti (JFN)

25

- *PRIMERI:*
  - devijacija date mere od prosečnog broja poena na testu iznosi -3
    - ne znamo: sam broj poena
    - ali: znamo da je taj broj *manji* (za 3) od proseka, tj. uspeh je ispodprosečan
  - devijacija visine ženske osobe od prosečne visine iznosi +10cm
    - ne znamo: koliko je osoba zaista visoka
    - ali: znamo da je *viša* (za 10) od proseka, dakle, njena visina je *natprosečna*
- utvrđivanjem *devijacije,* veličina *date* mere se postavlja u *kontekst* svih *drugih* mera u skupu
  - naime, *data* mera Y se poredi sa *karakterističnom* merom skupa, tj. prosekom M
  - matematički, to poređenje se vrši računanjem *za koliko* data mera odstupa od proseka, tj. računanjem *razlike:* $\boxed{d = Y - M}$
- sad se javlja sledeći problem: kada znamo devijaciju, još uvek ne znamo *odnos* te devijacije prema drugim *devijacijama*
  - *PRIMERI:* da li je data devijacija, npr. -3 ili +10, velika ili mala *devijacija*, u poređenju sa *drugim* devijacijama?
    - da li je uspeh na testu čija devijacija iznosi -3 samo malo ili veoma slabiji od prosečnog uspeha?
    - da li je osoba sa visinom 170cm izrazito visoka ili neznatno viša od proseka?

1. Jednofaktorski nacrti (JFN)

26

- pitanje: kako postaviti *datu* devijaciju d u *kontekst* svih *drugih* devijacija?
- odgovor: poređenjem *date* devijacije d sa *karakterističnom devijacijom* za ceo skup, a to je *standardna devijacija* SD
- matematički, poređenje se vrši utvrđivanjem *koliko puta* je d veća ili manja od SD, računanjem *količnika*: $\boxed{z = d / SD}$
- rezultat se naziva standard(izova)na mera ili z-skor, u oznaci z
  - veličina z stavlja datu devijaciju d u kontekst ostalih devijacija

PRIMER:

	mere	devijacije	z-skorovi
1.	165	+5	+5/7.1 = + 0.70
2.	150	-10	-10/7.1 = - 1.41
3.	160	0	0/7.1 = 0
4.	170	+10	10/7.1 = + 1.41
5.	155	-5	-5/7.1 = - 0.70
M	160	0	0

u ovom primeru: SD = 7.1

- uočimo: kao i d, i z-skorovi se moraju sabirati do nule
- (naime: $\Sigma z = \Sigma d/SD = 0/SD = 0$)

odnos veličina d i z

devijacije	z-skorovi
d = SD	z = 1
d > SD	z > 1
d < SD	z < 1

z = 1.41: data devijacija d=10 je 1.41 puta veća od SD=7.1

z = 0.7: data devijacija d=5 iznosi 0.7 od SD=7.1

predznak d i z

devijacije	z-skorovi
d = 0	z = 0
d > 0	z > 0
d < 0	z < 0

predznak od z govori o *smeru* razlike mere od proseka, na isti način kao i kod d

1. Jednofaktorski nacrti (JFN)

27

- z-skorovi (z = d / SD) se mogu računati za bilo koji skup mera, ali:
- ako mere imaju *normalnu distribuciju*, z-skorovi nose dodatnu informaciju
  - oni nam govore o *proporciji* odn. *procentu* mera koje se nalaze ispod, iznad, ili između nekih *specijalnih* z-skorova
- naime, ako su mere normalno distribuirane onda važi:
  - oko 68% (tj. oko 2/3) svih mera imaće z-skorove između z = -1 i z = +1
    - i to: polovina (34%) između -1 i 0, i polovina (34%) između 0 i +1
      - naime, normalna kriva je simetrična
  - oko 95% svih mera imaće z-skorove između z= -2 i z = +2
  - oko 99.7% svih mera (dakle skoro sve) imaće z-skorove između z= -3 i z = +3
    - ove činjenice se dokazuju u statističkoj teoriji
  - *PRIMER:* ako je za visine žena M=160 a SD=7, onda:
    - za visinu 167 važi da je d = +7, što znači da je z = d/SD = 7/7 = +1
    - za visinu 153 važi da je d = -7, što znači da je z = d/SD = -7/7 = -1
    - znamo da su visine približno normalno raspoređene u populaciji
    - zaključak: oko 2/3 svih žena imaće visine između 153 cm i 167 cm
    - oko 1/6 žena imaće visine preko 167 cm, a oko 1/6 visine manje od 153 cm
      - uočimo: ovo su veoma važni praktični zaključci za industriju odeće!

1. Jednofaktorski nacrti (JFN)

28

- odnos individualnih mera Y i z-skorova
  - setimo se: iz devijacione jednačine d = Y - M, izvodi se strukturna jednačina Y = M + d
  - uočimo: iz jednačine za z-skorove z = d / SD, može se izvesti jednačina d = z*SD
  - kombinacijom gornje dve jednačine dobijamo novu jednačinu:
    - Y = M + d = M + z*SD
- odavde sledi:
  - ako znamo *z-skorove* z nekog skupa mera, možemo, pomoću ove jednačine, na osnovu njih potpuno rekonstruisati i same te *mere* Y
    - uslov: poznati su nam prosek M i standardna devijacija SD skupa
- *PRIMER:*
  - neka z-skor visine neke osobe iznosi z = +2
  - neka je u datom skupu mera prosek visina M=160, a standardna devijacija SD=7
  - kolika je visina te osobe Y?
  - rešenje: Y = M + z*SD = 160 + 2*7 = 174 cm

1. Jednofaktorski nacrti (JFN)

29

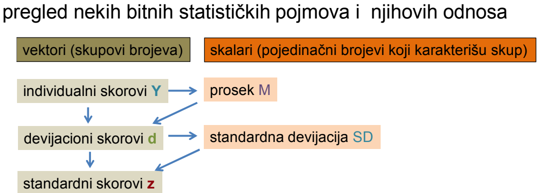
neke osobine SD i z-skorova

- u skupu z-skorova, SD može da služi kao *merna jedinica*
  - npr., kaže se da je visina neke osobe za 1 SD viša, a visina druge osobe za 1.5 SD niža od proseka
  - to je slično kao kad se, u skupu nestandardizovanih mera, kaže da je visina neke osobe za 10 cm viša a druge za 15 cm niža od proseka (ako je SD = 10 cm)
- ali, za razliku od uobičajenih mernih jedinica (kao što su cm, kg, itd), čije su veličine *fiksirane*, SD je *fleksibilna*
  - naime, SD se *menja* u zavisnosti od varijabilnosti datog skupa mera!
  - npr., u jednom skupu biće SD = 10 cm, u drugom će biti SD = 5 cm, itd
- pomoću SD mogu se *porediti* devijacije istog objekta za različite varijable (npr. visina i težina) sa različitim mernim jedinicama
  - *PRIMER:* za skup osoba je Mvis=180cm, SDvis=5cm, Mtež=80 kg, SDtež=10kg
  - određena osoba iz tog skupa ima nadprosečnu visinu (190 cm) i težinu (100 kg)
  - kako *uporediti* njene devijacije po visini (d=10 cm) i po težini (d=20 kg)?
    - da li je osoba 'viša (od proseka) nego što je teža (od proseka)'?
  - poređenje se vrši preko SD: u ovom slučaju, obe individualne devijacije d iznose po +2 SD, tj. ova osoba u *jednakoj* meri odstupa od proseka po visini i po težini

1. Jednofaktorski nacrti (JFN)

30

- standardizacijom *različite* varijable svodimo na iste *razmere*
  - to olakšava poređenja, slično kao računanje procenata kod frekvenci
  - *PRIMER:* da li uspeh na studijama (varijabla Z) više zavisi od školskog uspeha (varijabla X) ili od uspeha na prijemnom ispitu (varijabla Y)?
    - problem: X i Y imaju različite mere i ne mogu se direktno uporediti
    - rešenje: obe varijable se pre analize standardizuju
- pregled nekih bitnih statističkih pojmova i njihovih odnosa



napomena: da bi se vizuelno lakše razlikovali vektori i skalari, u daljem tekstu će vektori biti prikazivani **masnim** slovima (**bold**), kao gore